Background

The Cancer Assessment Review Committee (CARC) evaluated the epidemiological and experimental animal studies available for glyphosate employing the agency's 2005 Guidelines for Carcinogen Risk Assessment. The cancer guidelines emphasize the importance of a weight of evidence (WOE) approach in reaching conclusions about the human carcinogenic potential of pesticides. This is accomplished in a single integrative step after assessing all of the individual lines of evidence. The different lines of evidence considered include: 1] tumor findings, or lack thereof, in humans (primarily epidemiological studies); 2] tumor findings, or lack thereof, in laboratory animals; 3] chemical and physical properties; 4]its structure-activity relationships (SARs) as compared with other carcinogenic agents; and 5] studies addressing potential carcinogenic processes and mode(s) of action (MOA), either *in vivo* or *in vitro*. Data from epidemiological studies are generally preferred for characterizing human cancer hazard and risk. However, all of the information discussed above could provide valuable insights into the possible mode(s) of action and likelihood of human cancer hazard and risk (USEPA, 2005).

1. **Is either a positive trend or a pair-wise comparison enough to consider a study positive, once the other characteristics are taken in to account?**

Statistical considerations. The main aim of statistical evaluation is to determine whether exposure to the test agent is associated with an increase of tumor development. Statistical analysis of a long-term study should be performed for each tumor type separately. The incidence of benign and malignant lesions of the same cell type, usually within a single tissue or organ, are considered separately but may be combined when scientifically defensible (McConnell et al., 1986). Trend tests and pairwise comparison tests are the recommended tests for determining whether chance, rather than a treatment-related effect, is a plausible explanation for an apparent increase in tumor incidence. A linear trend test, such as the Cochran-Armitage test (Snedecor and Cochran, 1967) or the Exact Test for Trend (citation), asks whether the results in all dose groups together show an increase in tumor incidence as dose increases. Specifically, a trend test evaluates the likelihood that a line, which fits the dose response dataset, would have a slope greater than zero. A pairwise comparison test such as the Fisher exact test (Fisher, 1950) asks whether the incidence of a tumor type in one dose group is increased over that of the control group. By convention, for both tests a statistically significant comparison is one for which p is less than 0.05 that the increased incidence is not due to chance. Significance in either kind of test is sufficient to reject the hypothesis that chance accounts for the result (USEPA, 2005).

A statistically significant response may or may not be biologically significant and vice versa. The selection of a significance level is a policy choice based on a trade-off between the risks of false positives and false negatives. A result with a significance level of greater or less than 5% (the most common significance level) is examined to see if the result confirms other scientific information. When the assessment departs from a simple 5% level, this should be highlighted in the risk characterization. A two-tailed test or a one-tailed statistical test can be used. In either case a rationale must be provided. Statistical power can affect the likelihood that a statistically significant result could reasonably be expected. This is especially important in studies or dose groups with small sample sizes or low dose rates. Reporting the statistical power can be useful for comparing and reconciling positive and negative results from different studies. Considerations of multiple comparisons should also be taken into account. Haseman (1983) analyzed typical animal bioassays that tested both sexes of two species and concluded that, because of multiple comparisons, a single tumor increase for a species-sex-site

combination that is statistically significant at the 1% level for common tumors or 5% for rare tumors corresponds to a 7–8% significance level for the study as a whole**. Therefore, animal bioassays presenting only one significant result that falls short of the 1% level for a common tumor should be treated with caution (USEPA, 2005).**

CARC Assessment: OPP has followed the methods of statistical analyses traditionally used by the National Cancer Institute (NCI)/National Toxicology Program (NTP) for interpretation of carcinogenicity study findings (Fears *et al.* 1977; Hasemen, 1977; Gart *et al.*, 1979; Chu *et al.* 1981): Tumor data from the carcinogenicity studies in mice and rats are analyzed using the Fisher Exact Test for pairwise comparisons and the Cochran-Armitage or the Fisher Trend Test for the trend (dose-response) test; a fatal-tumor analysis is used when the tumor of interest is determined to have been fatal to the animals; and Peto Prevalence test is run when there are significant survival disparities between the dosed groups and the control group. The statistical methods used are dependent upon the data set.

Typically, rodent carcinogenicity bioassays (usually 2 species, 2 sexes, and 4 dose groups) include large number of comparisons with at least 30+ different tissue types examined, many of which will have background tumor rates. In some cases these outcomes may have a statistically significant trend and/or pairwise test due to chance alone, given background variation (i.e., a false positive). As illustrated in the table below, the suitability of relying on a trend test alone is questionable when the incidence in controls is zero and only the high dose group is responding with no pairwise significance. In this hypothetical case, there is no pair-wise significance for the single tumor seen at the high dose with the Exact Test for Trend but a positive trend is seen with the Cochran-Armitage Trend test.

Dose (ppm)	0	500	1000	3000
Tumor incidences	0/60	0/60	0/60	**1/60**
Exact Test for Trend	P=0.2500	P=1.0000	P=1.00000	P=0.5000
Cochran-Armitage Trend	P=0.0495*			

A number of published studies that evaluated the statistical methods used in tumor analyses of the NCI/NTP studies showed that the overall false positive error associated with linear trend test is about two time larger (20-25%) than that associated with control-high dose pairwise comparison test (10-11%). The trend tests are so sensitive, the Food and Drug Administration (FDA) in 2001 revised their decision rule from a statistical significance of $p < 0.01$ to $p < 0.005$ for common tumors and $p < 0.05$ to $p < 0.025$ for uncommon tumors. (Hasemen, 1983; Hasemen, 1984; Hasemen, 1990; Lin and Rahman, 1998; Lin, 2000).

In the case of glyphosate, in one study in mice, hemangiosarcomas were seen in 4/45 high-dose males compared to zero in the controls resulting in a positive trend but no pairwise significance. This is analogous to the example given above where a single tumor at the high dose resulted in a positive trend because the control incidence was zero which drove the statistical outcome. The positive trend was seen with both methods (Exact Trend or Cochran-Armitage):

Glyphosate – Hemangiosarcomas				
Dose (mg/kg/day)	0	100	300	1000
Hemangiosarcomas	0/47	0/46	0/50	4/45
EPA: Exact Trend Test (P)	0.003*	1.00000	1.00000	0.5332

IARC: Cochran-Armitage Trend Test (P)	0.001*	Not Reported

With regard to the kidney tumors in male mice in another study, a positive trend was seen only in the Cochran-Armitage trend test but not with the Exact Trend test:

Glyphosate – Kidney Tumors				
Dose (mg/kg/day)	0	161	835	4945
Adenomas/Carcinomas	1/49	0/49	1/50	3/50
EPA: Exact Trend Test (P)	0.0648	1.0000 0	0.7576	0.3163
IARC: Cochran-Armitage Trend Test (P)	0.034*	Not Reported		

As shown above, different statistical methods will produce different results (i.e. significance levels) when they are applied to the same data. Therefore, the most critical role of peer reviewers like the CARC is to determine which of these statistical signals are indeed test article-related and which ones are due to chance or to other corroborative evidence. The CARC does this by looking at all the contextual information of biological relevance, and not simply deferring to a significant *p* value. While the analysis of the data may be strictly a statistical issue, the interpretation of the study is not.

The hypothesis that a trend test will be testing whether there is a linear trend with a slope greater than zero, whereas, a pairwise comparison will be comparing whether a treated group is significantly different from the controls.

In the case of glyphosate, there are half-dozen or so equivocal findings, which is to be expected for a carcinogenicity dataset this large (11 studies as opposed to the typical 2 studies required for a food-use chemical). Linear trend tests are generally more sensitive than pairwise comparisons. As discussed above, occurrence of a single tumor at the high dose compared to zero in the control can result in a positive trend (i.e, false positive). Similarly, if the tumor incidences from all 11 studies resulted in a similar number of outcomes with a statistically significant *negative* trend, would one argue that glyphosate was having a chemopreventive effect?

It is critical that the final interpretation of the carcinogenic potential of a chemical consider not only the statistical evaluations of available datasets, but also the biological plausibility of the results. Statistical decision rules should not be employed as a substitute for sound scientific judgment in the overall evaluation of these experiments. In accordance with the agency's guidelines, the CARC uses the statistical significance of a given tumor incidence observed in an animal study in the overall ''weight of evidence'' approach when assessing carcinogenicity. It is noted that P-values are objective facts and the proper use of these values in the light of other information to decide whether or not the test compound really is carcinogenic involves subjective judgement. Therefore, it has been CARC's standard operating procedure to consider other evidence such as: 1) relative survival rate of dosed and control groups; 2) whether the tumor was seen in the target organ; 3) the appropriateness of combining tumors of varying sites and histogenesis for evaluation; 4) presence of related pre-neoplastic lesions; 5) progression to malignancy; 6) positive dose-response relationship; 7) incidence of other lesions of similar cell lineage; 8) type of tumor (e.g. background rates, rare vs common); 9) time of tumor onset; 10) occurrence of the tumor in both sexes of both species; 11) use of historical control data; 12) consistency and reproducibility across studies; 13) evidence for mutagenicity; 14) mechanistic

data; and 15) structure-activity concerns.

Regardless of the statistics, the biological plausibility of the reported tumors in glyphosate studies with mice points towards chance. In the four studies among glyphosate treated animals, the total incidence of kidney tumors was 3 out of 800 (0.4%) and hemangiosarcomas was 5 of 800 (0.6%). Reproducibility of findings across studies constitutes one of the strongest arguments for causality, and if these tumor types were treatment-related, they would have been at higher rates and in more than one study. The overall weight of evidence, unequivocally, indicates that glyphosate non-carcinogenic in animals.

It is generally believed that trend tests are more sensitive and so from the IARC's "precautionary" assessment point of view, a trend test may be preferable for identifying cancer "hazard" as defined by IARC's Preamble. However, in a regulatory setting, it is critical that a "weight of evidence" approach, as stipulated in the agency's guideline, is followed. For over three decades, OPP has strictly adhered to the weight of evidence approach in its evaluation of the carcinogenic potential of a chemical. In the case of glyphosate

2. **What weight is given to epidemiologic evidence in assigning descriptors for cancer classification?**

The EPA Cancer Assessment Guidelines address how a chemical assessment should integrate multiple lines of evidence and who to weigh different types of evidence in reaching a determination about the potential of the chemical to cause cancer in humans. When there are sufficient data to make a determination on human carcinogenicity, the Guidelines identify four possible classifications. The classification and associated guidance for each is presented below, followed by an explanation of how the CARC evaluated the available datasets for glyphosate against that guidance.

"Carcinogenic to Humans"

This descriptor indicates strong evidence of human carcinogenicity. It covers different combinations of supporting evidence.

- This descriptor is appropriate when there is **convincing epidemiologic evidence of a causal association** between human exposure and cancer.

  CARC: although there is a weak indication in four of ten epidemiological studies suggesting an association between glyphosate exposure and NHL, the other six studies render the epidemiology dataset far from convincing.

- Exceptionally, this descriptor may be equally appropriate with a lesser weight of epidemiologic evidence that is strengthened by **other lines of evidence**. It can be used when **all of the following** conditions are met:
  - (a) there is **strong evidence of an association** between human exposure **and** either cancer or the key precursor events of the agent's mode of action but

not enough for a causal association, **and**

- (b) **there is extensive evidence of carcinogenicity in animals, and**
- (c) the mode(s) of carcinogenic action and associated key precursor events have been identified in animals, **and**
- (d) there is strong evidence that the key precursor events that precede the cancer response in animals are anticipated to occur in humans and progress to tumors, based on available biological information.

CARC Assessment: In the case of glyphosate, this descriptor/classification is not applicable. The epidemiological evidence at this time does not support a causal relationship between glyphosate exposure and solid tumors or non-solid tumors (leukemia, multiple myeloma or Hodgkin lymphoma); epidemiolocal evidence at this time is inconclusive for a causal or clear associative relationship between glyphosate exposure and non-Hodgkin lymphoma (NHL). So the first criterion is not met. The evidence in experimental animals is far from extensive; there is no evidence for carcinogenicity following dietary administration for up to two years to male and female CD-1 mice (4 studies) and to Sprague-Dawley or Wistar rats (7 studies). There is no suggestion, much less strong evidence, that glyphosate's mode of action leads to a carcinogenic response. So, the second criterion is not met.

"Likely to Be Carcinogenic to Humans"

This descriptor is applicable when the weight of the evidence is adequate to demonstrate carcinogenic potential to humans but does not reach the weight of evidence for the descriptor "Carcinogenic to Humans". Supporting data for this descriptor may include:

- an agent demonstrating **a plausible (but not definitively causal**) association between human exposure and cancer, in most cases **with some supporting biological, experimental evidence,** though not necessarily carcinogenicity data from animal experiments;
- an agent that has tested positive in animal experiments in more than one species, sex, strain, site, or exposure route, with or without evidence of carcinogenicity in humans;
- a positive tumor study that raises additional biological concerns beyond that of a statistically significant result**,** for example, a high degree of malignancy, or an early age at onset;
- a rare animal tumor response in a single experiment that is assumed to be relevant to humans; or
- a positive tumor study that is strengthened by other lines of evidence**,** for example, either plausible (but not definitively causal) association between human exposure and cancer or evidence that the agent or an important metabolite causes events generally known to be associated with tumor formation (such as DNA reactivity or effects on cell growth control) likely to be related to the tumor response

CARC Assessment: In the case of glyphosate, this descriptor/classification is not applicable based

on the following WOE considerations:

A total of 10 case-control studies of glyphosate and NHL were evaluated. No statistically significant association between glyphosate exposure and NHL was seen in four studies and no association was seen with two case-control studies or in the Agricultural Health Study (AHS) prospective cohort study. Of the remaining four studies that showed a suggestive (but not causal) association, only one study showed a statistically significantly increased risk for NHL. However, there is not biological, experimental evidence to support a link between glyphosate exposure and NHL. The extensive experimental data (carcinogenicity, genotoxicity and/or mechanistic evidence) do not provide evidence to establish a basis for an association between exposure to glyphosate and the development of NHL, since glyphosate has not produced [any effects on the immune system or blood tumors?].

No treatment-related tumors were identified in 11 carcinogenicity studies. There is no evidence of carcinogenicity in 7 rat studies or two of the mouse studies. Kidney tumors were seen in male mice in one study and hemangiosarcomas were seen in male mice in another study. The increased incidence in these tumors, however, did not reach statistical significance in a pair-wise comparison with the concurrent controls. The presence of kidney tumors did not raise a biological concern since there were no pre-neoplastic lesions for the kidney; there was no high degree of malignancy or early onset of these tumors; and the tumors were only seen at a dose 5-fold higher than the limit dose. With respect to the hemangiosarcomas, hemangiosarcomas are commonly seen in this strain of mice, especially in males; and, the tumors were seen only at the limit dose. More importantly, neither of these tumor types was seen in the other three studies in the same strain of mice, demonstrating lack of reproducibility.

"Suggestive Evidence of Carcinogenic Potential"

This descriptor is appropriate when a concern for potential carcinogenic effects in humans is raised, but the data are judged not sufficient for a stronger conclusion. This descriptor covers a spectrum of evidence associated with varying levels of concern for carcinogenicity, ranging from a positive cancer result in the only study on an agent to a single positive cancer result in an extensive database that includes negative studies in other species. Depending on the extent of the database, additional studies may or may not provide further insights. Some examples include:

- a small, and possibly not statistically significant, increase in tumor incidence observed in a single animal or human study that does not reach the weight of evidence for the descriptor "Likely to Be Carcinogenic to Humans." The study generally would not be contradicted by other studies of equal quality in the same population group or experimental system;

- a small increase in a tumor with a high background rate in that sex and strain, when there is some but insufficient evidence that the observed tumors may be due to intrinsic factors that cause background tumors and not due to the agent being assessed. When there is a high background rate of a specific tumor in animals of a particular sex and strain, then there may be biological factors operating independently of the agent being assessed that could be responsible for the development of the observed tumors. In this case, the reasons for determining that the tumors are not due to the agent are explained;

- evidence of a positive response in a study whose power, design, or conduct limits the ability to

draw a confident conclusion (but does not make the study fatally flawed), but where the carcinogenic potential is strengthened by other lines of evidence (such as structure-activity relationships); or

- a statistically significant increase at one dose only, but no significant response at the other doses and no overall trend.

CARC Assessment: In the case of glyphosate, this descriptor/classification is not applicable based on the following WOE considerations:

A concern for potential carcinogenic effects in humans is raised; however, the totality of data from the epidemiological studies are judged not sufficient for a stronger conclusion since the epidemiological evidence at this time is inconclusive for a causal or clear associative relationship between glyphosate and NHL.

The weight of evidence from the carcinogenicity studies in mice and rats did not reach the level of evidence needed for either a "likely" or "suggestive" classification for the following reasons: The presence of kidney tumors in male mice in one study was contradicted by the lack of formation of this tumor type in the other three studies of equal quality in the same strain of mice. Similarly, the increased incidence of hemangiosarcomas observed in male mice in one study may also be due to an inherent increase in the background incidences and not due to treatment since this tumor type was not seen in the other three studies in the same sex and strain of mice when tested at comparable or even at high doses (4000 mg/kg/day). There was no structure-activity relationship evidence to support a carcinogenic potential since no evidence of carcinogenicity was seen in mice or rats administered sulfosate (the trimethylsulfonium salt of glyphosate) for two years.

"Not Likely to Be Carcinogenic to Humans"

This descriptor is appropriate when the available data are considered robust for deciding that there is no basis for human hazard concern. In some instances, there can be positive results in experimental animals when there is strong, consistent evidence that each mode of action in experimental animals does not operate in humans. In other cases, there can be convincing evidence in both humans and animals that the agent is not carcinogenic. The judgment may be based on data such as:

- animal evidence that demonstrates lack of carcinogenic effect in both sexes in well-designed and well-conducted studies in at least two appropriate animal species (in the absence of other animal or human data suggesting a potential for cancer effects),

- convincing and extensive experimental evidence showing that the only carcinogenic effects observed in animals are not relevant to humans,

- convincing evidence that carcinogenic effects are not likely by a particular exposure route, or
- convincing evidence that carcinogenic effects are not likely below a defined dose range.

CARC Assessment: In the case of glyphosate, this descriptor/classification is applicable based on the following WOE considerations:

Overall evidence from epidemiological studies are inconclusive for a causal associative relationship between glyphosate exposure and cancer in human studies. There is no evidence to support a causal relationship between glyphosate exposure and solid tumors and non-solid tumors (leukemia, multiple myeloma, or Hodgkin lymphoma). The evidence at this time is inconclusive for a causal or clear associative relationship between glyphosate and NHL. Multiple case-control studies and one prospective cohort study found no association; whereas, results from a small number of case-control studies (mostly in Sweden) did suggest an association. Limitations for most of these studies include small sample size, limited power, risk ratios with large confidence intervals, and recall bias as well as missing data. The literature will continue to be monitored for studies related to glyphosate and risk of NHL.

In experimental animals, there were no statistically significant increase in the occurrence of any tumor type in mice (4 studies) or rats (7 studies). The small increased incidences of kidney tumors in male mice in one study and hemangiosarcomas in male mice in another study were determined NOT to be treatment-related due to: the lack of pre-neoplastic changes in the kidneys; kidney tumors are unilateral with no evidence of multiplicity of form; hemangiosarcomas have a high spontaneous background levels in this strain of mice; lack of statistical significance in pair-wise comparison tests; lack of consistency in multiple studies; and the tumor incidences were within the historical control range. Furthermore, the lack of a dose-response across several orders of magnitude in multiple studies suggests that no individual tumor of a single etiology is attributed to treatment. Thus, these factors minimize the statistical significance seen in trend tests (but not in pair-wise comparison) *per se.* Additionally, a wide range of assays both *in vitro* and *in vivo* including endpoints for gene mutation, chromosomal damage, DNA damage and repair demonstrated that there is no *in vivo* genotoxic or mutagenic concern for glyphosate.

Chu, K, C, Cueto, K.C and Ward, J.M (198 1). Factors in the evaluation of 200 National Cancer Institute Carcinogen bioassays. J. Toxicol. Environ. Health 8, 25 I-280.

Fears, T. R., Tarone, R. E., and Chu, K. C. (1977).False-positive and false-negative rates for carcinogenicity screens. Cancer Res. 37, 194 1 - 1945.

Gart, J. J., Chu, K. C., and Tarone. R. E. (1979). Statistical issues in interpretation of chronic bioassays for carcinogenicity. J. Natl. Cancer Inst. 62,957-974.

Haseman, J.K (1977). Response to "Use of Statistics When Examining Lifetime Studies in Rodents to Detect Carcinogenicity". J. Toxicol.Environ.Health 3:633-636.

Haseman J. K. (1983). A reexamination of false-positive rates for carcinogenicity studies. Fundam. Appl.Toxicol. 12,793-804.

Haseman, J.K. (1984) Statistical Issues in the design, analysis and interpretation of animal carcinogenicity studies. Environmental Health Perspective 58 385-392.

Haseman, J.K., Hajian, G. Crump, K.S. Selwyn, M.R.& Peace, K.E. (1990) Dual control groups in rodent carcinogenicity studies. In Statistical Issues in Drug Research and Development, Ed, Peace, K. E. Marcel Dekker, New York.

Haseman, J.K., Huff, J. & Boorman, G.A (1984) Use of historical control data in carcinogenicity studies in rodents. Toxicologic Pathology 12 126-135.

Lin, K.K.& Rahman, M.A (1998) Overall false positive rates in tests for linear trend in tumor incidence in animal carcinogenicity studies of new drugs (with discussions). Journal of Pharmaceutical Statistics 81 1-22.

Lin, K. K (2000). Carcinogenicity studies of pharmaceuticals in Encyclopedia of Biopharmaceuticals Statistics edited by S.C.Chow, Marcel Dekker, New York, 88-103.